Walker, P. D., Rowe, M. T., Winter, A. J., Weir, A. N. M., Akter, N., Wang, L., Race, P. R., Williams, C., Song, Z., Simpson, T. J., Willis, C. L., & Crump, M. P. (2020). A Priming Cassette Generates Hydroxylated Acyl Starter Units in Mupirocin and Thiomarinol Biosynthesis. *ACS Chemical Biology*, *2020*. https://doi.org/10.1021/acschembio.9b00969

A Priming Cassette Generates Hydroxylated Acyl Starter units in Mupirocin and Thiomarinol Biosynthesis.

Paul D. Walker[1], Matthew T. Rowe[1], Ashley J. Winter[1], Angus NM Weir[1], Nahida Akter[1], Luoyi Wang[1], Paul R. Race[2], Christopher Williams[1], Zhongshu Song[1], Thomas J. Simpson[1], Christine L Willis[1,*], Matthew P. Crump[1,*]

[1] School of Chemistry, University of Bristol, Cantock's Close, BS8 1TS, Bristol, UK.

[2] School of Biochemistry, University of Bristol, University Walk, BS8 1TD, Bristol, UK.

ABSTRACT: Mupirocin, a commercially available antibiotic produced by *Pseudomonas fluorescens* and thiomarinol, isolated from the marine bacterium *Pseudoalteromonas* sp. SANK 73390, both consist of a polyketide-derived monic acid homologue esterified with either 9-hydroxynonanoic acid (mupirocin, 9HN) or 8-hydroxyoctanoic acid (thiomarinol, 8HO). The mechanisms of formation of these deceptively simple 9HN and 8HO fatty acid moieties in *mup* and *tml* respectively remain unresolved. To define starter unit generation the purified mupirocin proteins MupQ, MupS and MacpD and their thiomarinol equivalents (TmlQ, TmlS and TacpD) have been expressed and shown to convert malonyl-coenzyme A (CoA) and succinyl-CoA to 3-hydroxypropionoyl (3-HP) or 4-hydroxybutyryl (4-HB) fatty acid starter units respectively via the MupQ/TmlQ catalysed generation of an unusual bis-CoA/acyl carrier protein (ACP) thioester, followed by MupS/TmlS catalysed reduction. Mix and match experiments show MupQ/TmlQ to be highly selective for the correct CoA, MacpD/TacpD were interchangeable but an alternate *trans*-acting ACPs from the mupirocin pathway (MacpA/TacpA) or a heterologous ACP (BatA) were non-functional. MupS and TmlS selectivity was more varied and these reductases differed in their substrate and ACP selectivity. The solution structure of MacpD determined by NMR revealed a *C*-terminal extension with partial helical character that has been shown to be important for maintaining high titres of mupirocin. We generated a truncated MacpD construct, MacpD_T, which lacks this C-terminal extension but retains an ability to generate 3-HP with MupS and MupQ, suggesting further downstream roles in protein-protein interactions for this region of the ACP.

Mupirocin is a commercially available antibiotic marketed as Bactroban™. It is produced by *Pseudomonas fluorescens* used for the treatment of topical skin infections and as a pre-operative nasal spray. Mupirocin, which consists of a mixture of pseudomonic acids (PAs) A-C[1-4] is active against Gram-positive bacteria and most significantly is used clinically to treat methicillin resistant S*taphylococcus aureus* (MRSA).[5] The closely related thiomarinols (A-G) isolated from the marine bacterium *Pseudoalteromonas* sp. SANK 73390 combine scaffolds that are homologous to the pseudomonic acids with the addition of a pyrrothine moiety belonging to the holomycin class of antibiotics.[6-8] Both structures consist of a polyketide-derived monic acid homolog esterified with either 9-hydroxynonanoic acid (9HN, mupirocin) or 8-hydroxyoctanoic acid (9-HO, thiomarinol).

Mupirocin and thiomarinol are assembled by a type I *trans*-AT modular polyketide and fatty acid synthases (PKS/FAS).[9] The mupirocin synthase is encoded for by a 74 kb gene cluster consisting of 6 large open reading frames (ORFs), labelled mupirocin multifunctional proteins (MmpA to MmpF) and 29 ORFs (*mupA-X* and five trans-acting ACPs, *macpA-E*) encoding a range of tailoring enzymes.[10] MmpD and MmpA encompass the principle modular PKS components supplemented by smaller modules MmpC, MmpE, MmpF/MmpB that are thought to incorporate acyl transferase (AT)/ enoyl reductase (ER) activities, C10 - C11 epoxidation and fatty acid biosynthesis respectively. The entire thiomarinol cluster, interestingly encoded on a single plasmid, shares many of these common elements but lacks certain functions.[11] For example, the TmpE oxidoreductase domain appears to be inactivated whereas the equivalent mupirocin MmpE contains an oxidoreductase domain that appears to be essential for epoxidation of the C10 - C11 double bond.[12] Thiomarinol contains a C4 hydroxylation that is absent in mupirocin and it possesses the enzymatic machinery for addition of the pyrrothine moiety (**Figure 1**).

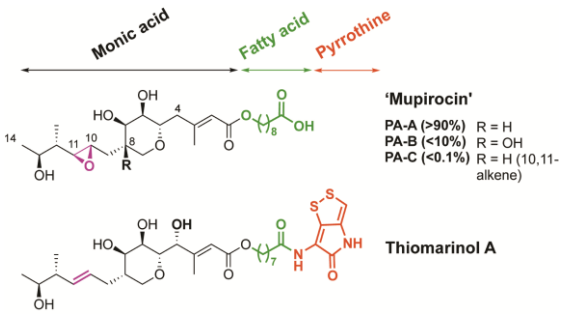


Figure 1. Mupirocin is a mixture of pseudomonic acids A-C. The structures are composed of a PKS-derived monic acid portion esterified to a fatty acid-derived 9-hydroxynonanoic acid. Thiomarinol A isolated from *Pseudoalteromonas* sp. SANK 73390, includes a pyrrothine moiety, a shorter 8-carbon fatty acid chain, C4 hydroxylation, no modification of C8 and lacks the C10-C11 epoxide (resembling PA-C). The minor metabolites thiomarinol B-G are not shown.

A conserved structural feature between mupirocin and thiomarinol A is the esterification of a polyketide-derived monic acid portion with a fatty acid. Feeding experiments with ^{18}O, ^{13}C and ^{2}H labelled acetates to *P. fluorescens* generated incorporation patterns consistent with the head-to-tail condensation of acetate in the polyketide portion of both mupirocin and 9HN (**Figure 2A, B**). To generate the odd number of carbon atoms, the 9HN fatty acid portion appeared to be derived from a 3-carbon starter unit incorporating a reversed acetate moiety that subsequently undergoes three acetate condensations to generate the C_9 acid.[13, 14] Similarly, a reduced level of labelled acetate incorporation was observed for C5'-C8' of the fatty acid chain of thiomarinol suggesting this portion of the chain may be derived from a 4-carbon starter unit. Feeding [2,3-$^{13}C_2$]-succinate or [2,3-$^{13}C_2$]-4-hydroxybutyrate gave ^{13}C-^{13}C coupling between C6' and C7' showing both were incorporated intact suggesting that 4-hydroxybutyrate may be derived through reduction of a succinate or succinyl-CoA (**Figure 2C, D**).[15]

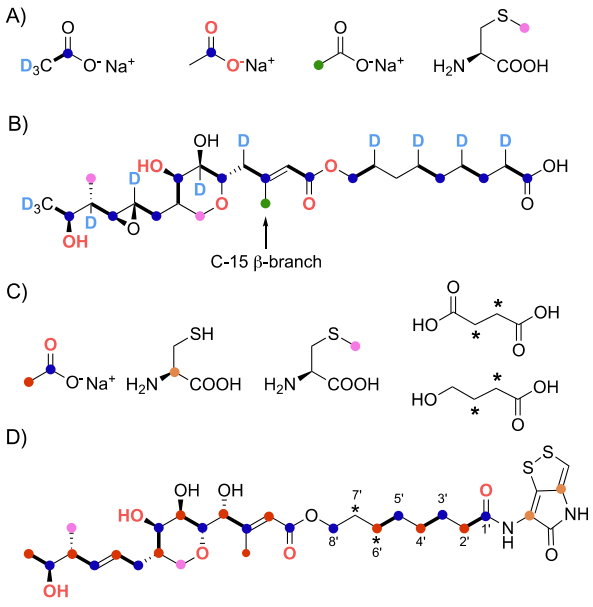


FIGURE 2. Isotopic feeding studies in the biosynthesis of mupirocin and thiomarinol. A) The acetates with differential isotopic enrichment fed to the mupirocin producing strain *P. fluorescens* NCIMB 10586 and B) The resulting labelling pattern of PA-A.[13, 14] C) Isotopically enriched precursors used in feeding studies with the thiomarinol WT producing strain *Pseudoalteromonas* sp. SANK 73390 and D) The resulting labelling pattern of thiomarinol A where colored dots correlate the positions of incorporation with the source of the [^{13}C] label. The asterisks (*) show the intact incorporation of 4-hydroxybutyrate, arising from succinate, into the fatty acid chain.[15]

The starter unit in mupirocin is therefore proposed to be a 3-hydroxypropionate derived moiety, formed by reduction of malonate. This undergoes three rounds of fatty acid biosynthesis to give a C_9 chain, whereas thiomarinol presumably primes with 4-hydroxybutyrate derived moiety that undergoes two elongations to give a C_8 chain. In support of this priming, different fatty acid chain lengths have been isolated in strains lacking one or more of the enzymes involved in polyketide chain processing, *e.g.* mupirocin W, desepoxymupirocin W and marinolic acids which show fatty acid lengths of C_4 to C_8.[12, 15] To date, the mechanism for the formation of both of these starter units is unproven. The most likely origin has been hypothesized to be a three-enzyme cassette consisting of an ACP (MacpD/TacpD), MupQ/TmlQ and MupS/TmlS. Sequence homology suggests MupQ/TmlQ belong to the class I adenylate forming superfamily (ANL family) of enzymes that includes acyl-CoA synthetases, NRPS adenylation domains (A-domains), luciferases, fatty acid-AMP ligases (FAALs) and the topologically distinct pimeloyl-CoA synthetase (Supporting Information Figure S1).[16-19] Similarly, sequence homology has suggested MupS/TmlS to possess 3-oxoacyl ACP reductase functions. Mup and Tml enzymes show significant sequence identity (45.4%, 46.1% and 61.8% for the ACPs, MupQ/TmlQ and MupS/TmlS respectively) and functional homology to a cassette of enzymes (DifC, DifD, DifE) from the difficidin biosynthetic pathway from *Bacillus amyloliquefaciens*. These enzymes were proposed to form the acryloyl thiol ester starter unit from reduction of pyruvate (hypothesized to be DifE) and a subsequent dehydration. A role for DifD in transferring the acryloyl thiol ester from the putative ACP DifC to the N-terminal priming domain of the PKS was also proposed.[5, 20] To date the evidence for this mechanism has remained circumstantial and *in vivo* knock-out experiments (ΔMupQ, ΔMupS, ΔMacpD) have yielded the early, but otherwise uninformative shunt products, mupiric acid and mupirocin H.[21, 22] More generally we have observed knock-outs that generate this phenotype seem to be associated with late or post monic acid assembly.

In this study we definitively prove the mechanisms for the generation of the starter units from both the mupirocin and thiomarinol biosynthetic pathways by *in vitro* reconstitution of these steps using purified enzymes. The three and four-carbon starter units are derived from malonyl- and succinyl-CoA respectively and proceed via the reduction of an unusual bis-CoA/ACP thioester. The presence of the ACP is essential and MupQ/TmlQ perform the role of gatekeepers that select for the correct acyl CoA. In addition, we demonstrate by NMR that MacpD contains an extended C-terminus with a propensity to form an additional α-helix when compared to other known ACPs. Although important

for function this is not required for its activity in starter unit formation, suggesting a role in downstream interactions.

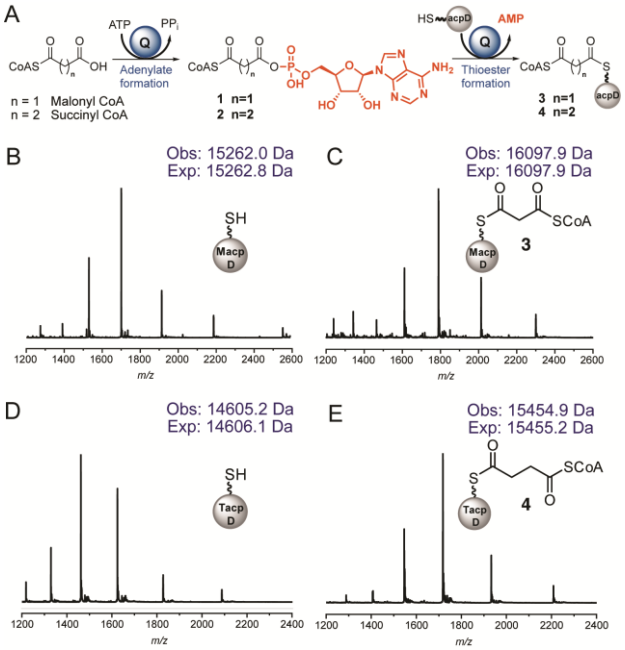
RESULTS AND DISCUSSION.

MupQ and TmlQ are adenylation domains. The homology of MupQ and TmlQ to the ANL family (Supporting Information Figure S1) suggested these enzymes might select malonyl-CoA and succinyl-CoA respectively and activate them to the adenylates **1** (n =1) and **2** (n=2) respectively, prior to attack by *holo*-MacpD or TacpD, forming ACP derivatives **3** and **4 (Figure 3A).**[17,23] MacpD, TacpD, MupQ, MupS, TmlQ and TmlS were cloned, recombinantly over-expressed and purified to homogeneity (**Supporting information Figure S2-S4, supporting information Tables 1 and 2**). Recombinant MacpD/TacpD were both over-expressed in the apo form and subsequently converted to the *holo* form using MupN.[24]

Holo-MacpD (**Figure 3B**) was incubated with MupQ, malonyl CoA and ATP. Electrospray mass spectrometry (ESMS) analysis showed consumption of *holo*-MacpD and confirmed quantitative formation of a new species with a mass consistent with **3** (observed 16097.9 Da; expected 16097.9 Da), a three-carbon intermediate attached *via* thioester linkages to both MacpD and CoA (**Figure 3C**).

Removal of ATP from the assay gave no product, confirming that it was essential for the reaction and MacpD was unable to self-malonylate.[25] The requirement for ATP was consistent with the formation of malonyl adenylate **1** (n=1) although this intermediate was not isolated. Denaturing MupQ prevented conversion to product confirming enzymatic activity (**Supporting information Figure S5**). The native thiomarinol system was also reconstituted *in vitro* by incubating TmlQ, ATP and succinyl-CoA with *holo*-TacpD which was converted to **4** with a mass of 15454.9 Da (Expected 15455.2 Da, **Figure 3D, E**). This was again consistent with formation of a TacpD/CoA bis-thioester but with an extended four-carbon chain.

Figure 3. Reaction of M/TacpD with native adenylation domains. The phosphopantetheine arm of the ACP is represented by a wavy line. A) Proposed reaction sequence for generation of bis-thioesters 3 and 4. Abbreviation Q denotes either MupQ or TmlQ. B) ESMS mass/charge (m/z) envelopes and deconvoluted mass calculation (Obs) and expected mass (Exp) for *holo*-MacpD. C) The reaction mixture of *holo*-MacpD + MupQ + malonyl-CoA. D) *Holo*-TacpD. E) The reaction mixture of *holo*-TacpD + TmlQ + succinyl-CoA.



MupS encodes a Malonyl-3-CoA-1-ACP reductase. Sequence homology predicted MupS to be a member of the NAD(P)H-dependent oxidoreductase family of enzymes. A deposited X-ray structure of MupS (PDB: 4yxf) without cofactor shows a classic Rossmann fold and a putative catalytic tetrad (Asn125, Ser153, Tyr166 and Lys170) that may be capable of two successive reductive steps from the CoA thioester to the alcohol (**Figure 4A**). The predicted reduction by MupS/TmlS was therefore investigated using the ACP/CoA bis-thioester product of MupQ/TmlQ. Addition of MupS and an excess of NADH to the 3-carbon malonyl intermediate **3** bound to MacpD resulted in the reduction to a species consistent with the formation of MacpD-3-HP **5** (**Figure 4B**). The observed ESMS trace of the derivatised MacpD could not however resolve either partial reduction to the malonic semi-aldehyde or further reduction to the alcohol. The product was therefore analysed via a phosphopantetheine ejection assay (Ppant ejection)[26] whereby collisional energy triggers ejection (fragmentation) of the derivatised phosphopantetheine arm from the ACP, allowing high accuracy mass determination of the resulting low molecular weight species. This analysis yielded a species of 333.14 Da (**Figure 4C**) which definitively proved generation of the 3-hydroxy-propianoyl group via two rounds of reduction. Similarly, the action of TmlS and NADH on **4** was monitored by ESMS and Ppant ejection assay which confirmed generation of TacpD-4-HB **6** (**Figure 4D, E**). Finally, repeating the assay with MacpD, MupQ, MupS and only one equivalent of NADH, an intermediate by Ppant ejection corresponding to the partially reduced malonic semi-aldehyde was detected (**Supporting information Figure S5**).

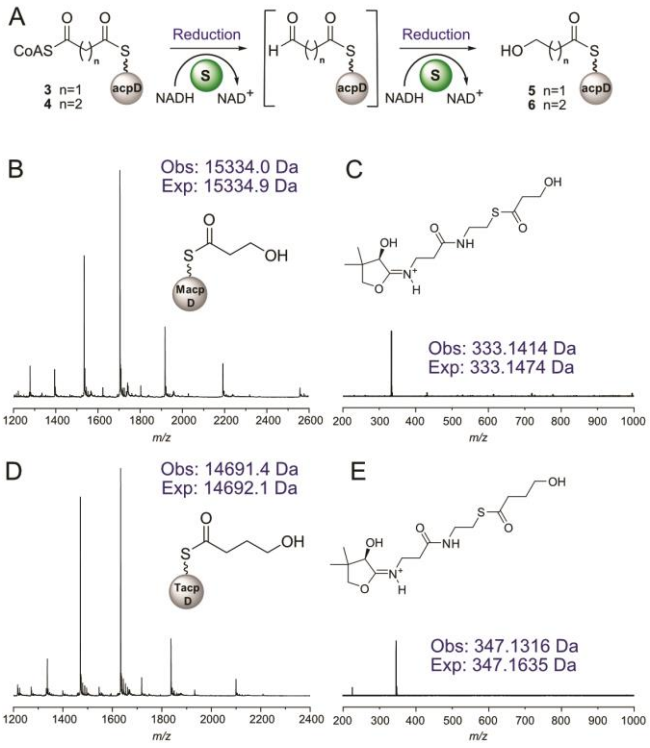


Figure 4. Reduction of 3- or 4- carbon ACP/CoA intermediates by native reductase domains. A) Proposed reaction sequence for reduction of 3 and 4 to yield MacpD-3-HP 5 and TacpD-4-HB 6 respectively. S denotes either MupS or TmlS. B) ESMS mass/charge (m/z) envelope and deconvoluted mass calculation (Obs) and expected mass (Exp) for MacpD + MupQ + malonyl-CoA + MupS + NADH and C) Corresponding Ppant ejection for B. D) Electrospray mass/charge (m/z) envelope for TacpD + TmlQ + succinyl-CoA + TmlS + NADH and E) Ppant ejection assay for D.

MupQ is selective for the priming CoA. Having shown the formation of the 3- or 4- carbon intermediates bound between ACP and CoA thioesters, the selectivity of the adenylation step was investigated by swapping the ACP, the adenylation domain and the CoA.

First only the CoA substrate was swapped and succinyl CoA was incubated with MacpD/MupQ and malonyl CoA with TacpD/TmlQ (**Table 1, entries 2 and 6**). Analysis by ESMS revealed that neither experiment yielded the reversed equivalents of the bisthioesters containing 4- or 3- carbons respectively. However, non-productive trans-thioesterification of the succinyl/malonyl group from CoA to *holo*-MacpD/TacpD was observed in both cases (data not shown).

Next, MacpD and TacpD were swapped. Incubation of MacpD with succinyl CoA and TmlQ resulted in the quantitative formation of a four-carbon activated intermediate with a mass of 16110 Da (**Table 1 entry 4, Supporting information Figure S6**). Similarly, when TacpD was incubated with malonyl CoA and MupQ, complete conversion to the respective 3-carbon intermediate was observed (**Table 1 entry 8, Supporting information Figure S6**). This demonstrated that both ACPs were compatible with non-native adenylation domains and the adenylation domains alone controlled substrate selectivity. Likewise, combinations of the correct ACP and substrate could not override the selectivity of a non-native adenylation domain and these combinations were inactive (**Table 1 entries 3 and 7**).

Entry	ACP	Co-factor	Q	Adenylation
1	MacpD	Mal	MupQ	Y
2		Succ	MupQ	N
3		Mal	TmlQ	N
4		Succ	TmlQ	Y
5	TacpD	Succ	TmlQ	Y
6		Mal	TmlQ	N
7		Succ	MupQ	N
8		Mal	MupQ	Y
9	MacpA	Mal	MupQ	N
10		Succ	TmlQ	N
11	TacpA	Succ	TmlQ	N
12		Mal	MupQ	N
13	BatA	Mal	MupQ	N

Table 1. Native and Mix and match experiments varying the ACP, substrate and adenylation domain (Q) and monitoring product formation by ESMS (Supporting Information Figure S6 and Figure S7). Successful reactions occurred for the correct co-factor/adenylation domain pair (MupQ + malonyl-CoA or TmlQ + succinyl-CoA) to introduce a 3- or 4-carbon unit respectively. M/TacpA and BatA were unable to complement for M/TacpD.

MacpD and TacpD are both essential for starter unit formation. In addition to MacpD and TacpD, both the mupirocin and thiomarinol synthases contain further free-standing ACPs (in addition to the Type I and tandem array ACPs), several of which have not been functionally characterised. These include MacpA and TacpA and previous work has shown that MacpA is unable to complement MacpD *in vivo*,[27] suggesting specific interactions for each ACP. To test *in vitro* whether lack of complementation was specifically due to loss of starter unit formation, MacpA was incubated with malonyl CoA/MupQ or succinyl CoA/TmlQ under standard conditions (**Table 1, entries 9 and 10**). In either case, no evidence of the desired bis-thioester intermediate was detected by MS. Similarly, TacpA was unable to substitute for TacpD (**Table 1, entries 11 and 12)**. The only products observed arose from trans-thioesterification of the CoA-bound acyl groups to the thiol of *holo*-MacpA or TacpA (**Supporting information Figure S7**). When malonyl-CoA was used, decarboxylation was observed leading to the generation of acetyl-MacpA/TacpA (**Supporting information Figure S7**). Further evidence for the high ACP-adenylation domain specificity was confirmed by the inability of BatA, a 4-helix, free-standing ACP from the β-branching cassette of the kalimantacin biosynthetic pathway,[28] to function with malonyl CoA/MupQ (**Table 1 entry 13**).

TmlS and MupS show different ACP and substrate specificities.

The selectivity of the reductase domains was investigated through further mix and matching of substrate and MacpD. MupS was able to reduce the correct 3-carbon substrate when paired with either MacpD or TacpD (**Supporting Information Figure S8**) but failed to act upon the 4-carbon substrate when combined with either ACP (**Table 2 entries 1 and 7**). Conversely, TmlS was only functional when paired with TacpD, irrespective of which substrate was present, but when paired with TacpD is was able to reduce both 3- and 4-carbon substrates (**Table 2 entries 5 and 8**).

Entry	ACP	Chain length	S	Reduction
1	MacpD	C3	MupS	Y
2			TmlS	N
3		C4	TmlS	N
4			MupS	N
5	TacpD	C4	TmlS	Y
6			MupS	N
7		C3	MupS	Y
8			TmlS	Y

Table 2. Reduction of 3- or 4- carbon bis-thioesters by native or non-native reductase domains (Supporting Information Figure S8).

The C-terminal MacpD extension is not essential for starter unit generation. MacpD is an unusually large ACP (106 AAs) with a predicted C-terminal extension. Previous complementation assays showed a truncated MacpD lacking the C-terminal twenty-one amino acids could partially rescue a ΔMacpD mutant but yielded a reduced titre of mupirocin, suggesting it may play a role in enhancing protein-protein interactions.[27] Similarly MacpA possesses a C-terminal extension although in this case the sequence was essential for mupirocin production.[27] Both of the equivalent thiomarinol ACPs also share these C-terminal extensions.

To further characterize MacpD we determined the three-dimensional structure of this protein using solution state NMR (**Figure 5**, **Supporting information Table 3 for structural statistics, Supporting information Figure S9**). The structure revealed the core of the protein is comprised of an archetypal ACP four-helix bundle, with the addition of partial helicity in the extended C-terminal domain (**Supporting information Figure S9**). Although this showed some propensity for folding, we did not identify any interactions to suggest this region associated with the core of the protein. The ACP presents a negatively charged/polar face along helices 2 and 3 and the C-terminus is a mixture of negatively charged and hydrophobic groups (**Figure 5**).

The presence of a partial C-terminal helix led us to investigate whether this region specifically influenced starter unit generation. We expressed and purified a truncated MacpD construct (MacpD_T), lacking fifteen C-terminal residues (**Supporting information Figure S10**). When the assays for starter unit generation were repeated using MacpD_T, we were able to demonstrate formation of 3-HP-MacpD (**Supporting information Figure S10**). Preparation of ^{15}N-labelled full length apo MacpD and titration with MupQ monitored by heteronuclear single quantum coherence spectra (HSQC)showed the most significant chemical shift perturbations (CSPs) located in helix 1, and line-broadening of Asp47 detected in helix 2 (**Figure 5C, D and G**). There were no discernible CSPs in the C-terminal region.

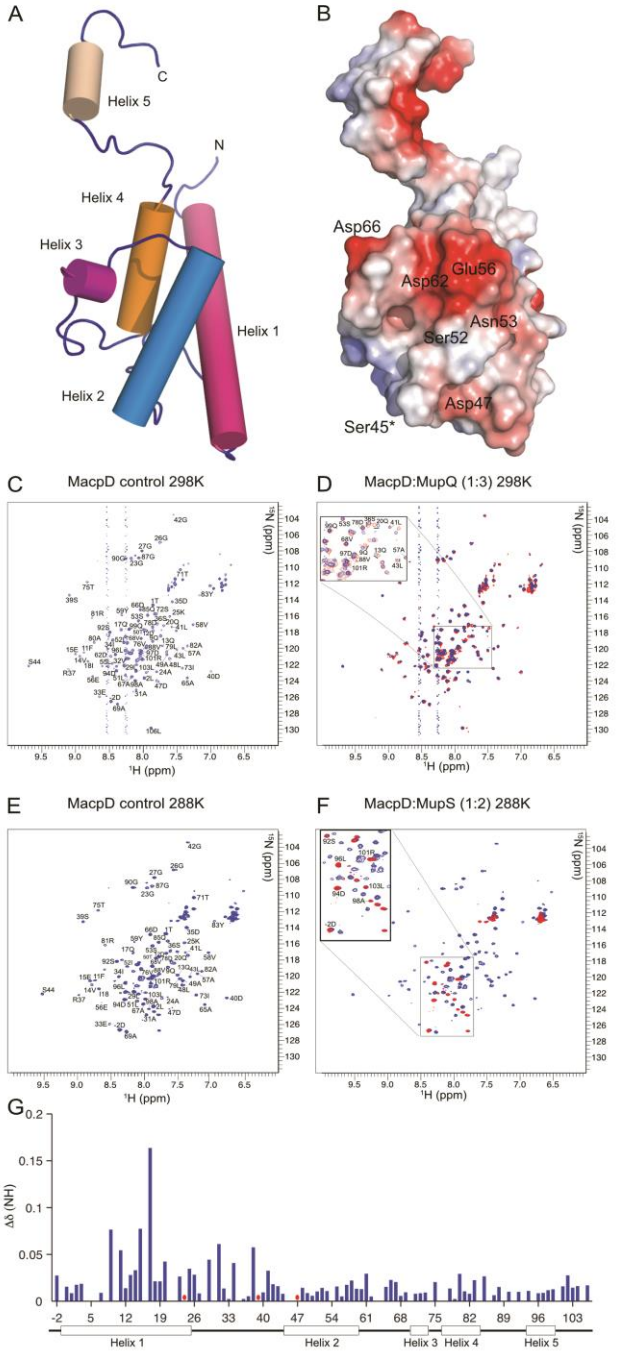
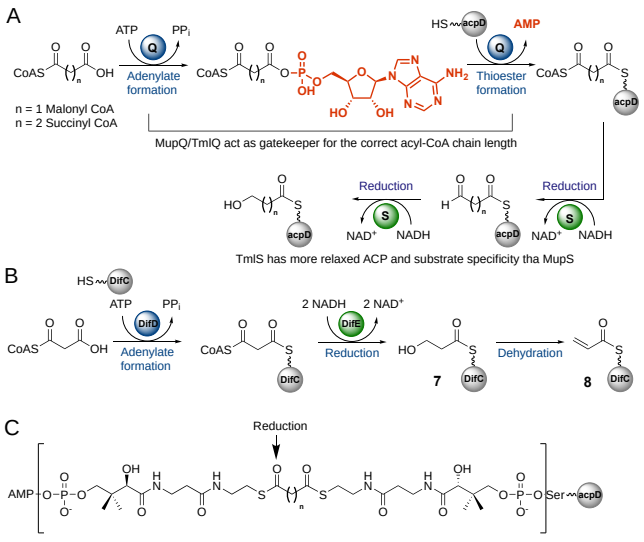


Figure 5. Three-dimensional solution structure of MacpD and ^{1}H-^{15}N HSQC experiments to probe protein-protein interactions between MacpD and MupQ/MupS. A) Representative model of MacpD with helices shown as cylinders and the short/partial helix labelled 3. B) Surface charge map showing predominantly negative and polar groups along helices 2 and 3 and the interconnecting loop (Asp62). Ser45 is the site of phosphopantetheinylation. C) Assigned control spectrum of free apo-MacpD at 298K and pH 8.0 (blue) and D) In the presence of three equivalents of MupQ (red). Small but discernible chemical shift

perturbations are observable. E) Apo-MacpD control at 288K, pH 8.0 and F) In the presence of two equivalents of MupS (red). The non-perturbed C-terminal resonances remain visible G) Chemical shift perturbations for apo-MacpD titrated with MupQ. Red dots indicate broadened residues and those with a CSP of 0 were either proline residues or not visible at pH 8.0.

Titration with MupS however gave striking changes, with almost all of the resonances of MacpD being broadened or vanishing, but with the clear exception of resonances originating from the C-terminal helix which remained visible (**Figure 5E and F**). Together, the functional assays and NMR titrations confirmed that the C-terminal helix is not involved in interactions with either MupQ or MupS.

MupQ selectivity and homology. Starter unit diversity can imbue specific biological properties on a natural product and provide the opportunity for pathway engineering as a route to novel metabolites. In a highly specific manner, malonyl CoA is selected and activated by MupQ, whilst succinyl CoA is selected by TmlQ (**Scheme 1A**). This selectivity or gatekeeper role correlates with the comparative role of NRPS adenylation domains that are responsible for activating the correct amino acid extender units and assembling the correct primary sequence.[29, 30] MupQ/TmlQ and NRPS A-domains belong to the ANL family that has been divided into nine groups[31] (of which acyl CoA synthetases and NRPSs are just two examples) that share a common two-step mechanism of carboxylate substrate adenylation and thioesterification. MupQ/TmlQ are rather unusual in that the substrate, malonyl- or succinyl-CoA, is already thioesterified but the free carboxylate undergoes adenylation and thioesterification with ACP to form a bis-thioester. As far as we are aware there is only one similar example of bis-thioester formation recently reported for the ambruticin biosynthetic pathway. Here the ACP of an unusual FAAL-ACP didomain enzyme, AmbG, is proposed to activate a modular (AmbF) ACP-bound polyketide via formation of bis-thioester, effectively cross-linking two PKS synthase components.[32]



SCHEME 1. Starter unit generation by adenylation, thioesterification and subsequent reduction. A) MacpD, MupQ and MupS produce a 3-hydroxypropionate whilst TacpD, TmlQ and TmlS produce 4-hydroxybutyrate. Abbreviations Q and S denotes MupQ/TmlQ and MupS/TmlS respectively. B) A homologous cassette of enzymes in the biosynthesis of difficidin produces a 3-carbon acrylate starter unit. C) Reduction of the ACP/CoA bis-thioester highlighting the selectivity required.

Homology modelling suggests MupQ/TmlQ share the common ANL family fold characterized by two domains, an N-terminal domain (~350 aa) and a smaller C-terminal domain (~100 aa), that undergo significant rearrangements through the catalytic cycle.[23] Conserved sequence motifs have been identified for the ANL superfamily in both the NRPS adenylation domains (A1-10)[29] and for aromatic adenylate forming ligases (Motif I-III).[33] MupQ contains, for example, the canonical sequence motifs I/A3, II/A5, III/A7/A8, A10 which are responsible for acting as the phosphate binding loop (I), the ATP binding site and Mg^{2+} coordination (II), coordination of the ribose ring of ATP (III), structural roles (A6) and acyl-CoA binding (A10). These encompass the 100% conserved residues Glu277 (Mg^{2+} binding), Gly319 (structural), Asp337/Arg351 (ATP binding) and Lys455 that are distributed at the ATP/Mg^{2+}/Acyl-CoA/4'-PP binding site in a three-dimensional model of MupQ generated by I-TASSER (**Supporting Information Figure S11A,C**).[34] The corresponding motifs are also present in TmlQ (**Supporting Information Figure S11**). Adjacent to these motifs acyl binding pockets from a range of substrates of differing sizes and chemistries have been characterized and correlated with the polarity and size of the acyl binding pocket. This pocket shows a number of conserved residues in MupQ and TmlQ as well as subtle differences which may combine to select either malonate or succinate as substrates (**Supporting Information Figure S11B**) but will await future structural and functional studies.[35] This pocket opens to a putative binding interface that is predominantly positively charged (residues Arg201, Arg232 and Lys379 in MupQ) that might provide a suitable binding site for the negatively charged residues likely to be presented by the helix 2 and 3 residues surrounding the phosphopantetheine arm of the ACP (*e.g.* Asp47, Glu56, Asp62, Asp66 in MacpD).

MupS selectivity and homology. The resulting three-carbon bis-thioester with CoA and MacpD is subsequently reduced by MupS and two equivalents of NADH to form MacpD-3-HP. In thiomarinol biosynthesis, *holo*-TacpD receives activated succinyl-CoA from TmlQ, and the resulting four-carbon intermediate is reduced by TmlS and NADH to form TacpD-4-HB. Whereas the adenylation domains (MupQ/TmlQ) were highly substrate selective, functional mismatches with MupS and TmlS were less predictable and a degree of promiscuity was observed. Biosynthetically this is logical as the adenylation domains have already completed their role as a gatekeeper for selection of a three- or four-carbon substrate.

MupS also shows homology to the SDR domain of the malonyl-CoA reductase (MCR), a bifunctional enzyme (alcohol and aldehyde dehydrogenase activities) from *Chloroflexus auranticus*. The MCR generates 3-hydroxypropanoic acid (3HP) used in CO_2 fixation from reduction of malonyl-CoA via the malonate semi-aldehyde.[36] In the related difficidin biosynthetic pathway, it is therefore likely that DifE operates in a similar fashion and generates **7** which is subsequently dehydrated to generate the acroyl-ACP starter unit **8** (**Scheme 1B**).

MupS and homologues evidently achieve both reductions using a single active site, binding a so-called 'bulky-bulky' thioester **(Scheme 1C**) where 'bulky' groups sit either side of the target group to be reduced as well as the bulky malonic-semialdehyde following one round of reduction. The pantetheine arm of an ACP is derived from CoA, so the bis-thioesters bound by MupS/TmlS show remarkable symmetry extending away from the centre of symmetry over 16 or 17 consecutively bonded atoms respectively and terminating in the phosphodiester bond to the conserved serine of helix 2 in the ACP or the adenine group of CoA. Presumably correct orientation of this intermediate is critically dependent on ACP/MupS recognition to ensure substrate delivery from the correct portal and prevent aberrant reduction of the ACP thioester group. Dual specificity has been observed previously using the Ras ADH from Ralstonia, which could reduce both bulky-bulky ketones as well as the corresponding thioester substrates (non ACP-bound). It was noted, however, that reduction of the bulkier thioester was on the order of ~500 times slower than the ketone.[37]

Efforts to use *in vitro* assays to delineate the non-linear arrangement of the enzymes of the mupirocin pathway have been hampered by an apparent over-abundance of type II ACPs including MacpA-E. MacpE has recently been confirmed to be involved only in the late stage processing of PA-B to PA-A,[38] MacpC is involved in β-branching of the polyketide backbone[24] and the current study has confirmed MacpD is involved in starter unit production. The role of the C-terminally extended MacpA is currently not defined.

Following starter unit generation, the next step of mupirocin biosynthesis is unknown. Once loaded with 3HP MacpD must interact with further partner proteins involved in either fatty acid extension of the starter unit, esterification with monic acid or even tetrahydrofuran formation, depending on the timing of these step.[12] Analysis of metabolite production by WT and mutant *P. fluorescens* strains have revealed pseudomonic acids with reduced fatty acid chain lengths and lacking the tetrahydrofuran ring, suggesting these are late stage modifications and esterification may precede these steps. The C-terminal helix has been shown to be important for MacpD function, and critical for MacpA, and further functional and biophysical binding assays should be able to delineate the wider role of these extended ACPs and the remaining *trans*-acting ACP, MacpB. Further candidate proteins for interaction partners include the ketosynthase domains of MmpF/MmpB which may co-operate to catalyse the carbon-chain extension of these starter units via the Claisen condensation of ACP-bound malonyl extender units that may require specific docking with MacpD and/or MacpA/B. Some, but perhaps not all, of the mechanistic steps involving an ACP may be directly transferable to the thiomarinol biosynthetic pathway.

MATERIALS AND METHODS

Reagents were purchased from Sigma Aldrich, Thermo Fisher or Merck Millipore. *E. coli* competent cells were purchased from New England Biolabs (T7 Express and 5-α) or Merck Millipore (Novagen BL21 (DE3)). All enzymes used were purchased from Thermo Fisher Scientific.

Plasmid Preparation. Details of sequences and primers used are given in the Supporting Information. The genes encoding M/TacpA, TacpD, MupQ, MupS, TmlQ, TmlS were amplified from genomic DNA and cloned into pOPINF[39] whilst the gene for MacpD and MacpD_T was synthesized (Thermo Fisher, GeneArt) and cloned into pET151. A C80A mutation was introduced into full-length MacpD for long term ACP stability and structural studies.

Expression and Purification of MupQ, TmlQ, MacpD, TacpD, MacpD_T, **MacpA and TacpA**. *E. coli* competent cells transformed with plasmid DNA were grown 37 °C for 16 h with shaking at 200 rpm. 2 mL of this seed broth was used to inoculate flasks containing 200 mL LB/carb medium and incubated at 37 °C with shaking at 200 rpm until an OD_{600} of 1.0 was reached. The culture was induced with 250 μM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated for 16 h at 16 °C. Cells were harvested by centrifugation (5000 x g) and re-suspended in column buffer A (50 mM Tris-HCl, 500 mM NaCl, 10% glycerol, pH 8.0). The cells were lysed by sonication and the soluble fraction applied to a HisTrap HP 5 mL Ni column (GE Life Sciences). Protein was eluted using a linear gradient from 5 to 100% column buffer B (50 mM Tris-HCl, 100 mM NaCl, 10% glycerol, 0.8 M imidazole, pH 8.0). 1 mL Fractions were exchanged into reaction buffer (50 mM Tris-HCl, 100 mM NaCl, pH 8.0) using a HiPrep 26/10 desalting column (GE Life Sciences) or further purified by size exclusion chromatography (Superdex S75) using reaction buffer. Protein was concentrated (maximum ~50 μM before precipitation for MupQ/TmlQ) and stored at 4 °C, or supplemented with 10% glycerol, aliquoted, flash frozen and stored at –80 °C.

Expression and Purification of MupS and TmlS. Both enzymes were prepared as described above with the exception that during purification 0.5 M NaCl and 10% glycerol was required in the reaction buffer to prevent precipitation.

Starter unit assay. Apo-ACP (~200 μM) was converted to *holo*-ACP by the addition of $MgCl_2$ (10 μM), CoA (1 mM), MupN (1.5 μM) and TCEP (1 μM). The reaction was monitored by MS and upon completion was desalted into reaction buffer supplemented with 1 mM TCEP using a HiPrep 26/10 desalting column. *Holo*-ACP was adjusted to ~100 μM, aliquoted and stored at – 20 °C. To *holo*-ACP (100 μM) were added MupQ/TmlQ (5 μM), malonyl/succinyl CoA (1 mM) and ATP (1 mM). The reaction was incubated at room temperature for 1 h then monitored by MS. MupS/TmlS (5 μM) and NADH (1 mM) were added and the reaction incubated for a further 1 h prior to MS analysis. For the detection of the partially reduced malonic semialdehyde, NADH was restricted to 0.1 mM. BatA, used as a control in these assays, was expressed and purified as described previously[40]

ESI-MS. Samples were desalted for ESI-MS analysis using a C_4 ZipTip™ (Millipore) per manufacturers instructions. Denatured samples were analysed on a Synapt G2-Si (Waters) fitted with a TriVersa NanoMate® (Advion) using the following parameters: sample cone 10 V, capillary voltage 1.5 kV, transfer collision energy 5 V. The source was set to positive mode and spectra were acquired over 600-3000 m/z and analysed using MassLynx™ 4.1 software. For Ppant ejection assays, an appropriate charge state was isolated using the MSMS functionality. The trap collision energy was increased until fragmentation was observed (typically 5 V – 30 V) and spectra were collected from 200-2000 *m/z*.

Expression and purification of ^{15}N and ^{13}C MacpD.

Single (^{15}N) and double labelled (^{15}N,^{13}C) samples of MacpD were prepared using M9 minimal media supplemented with NH_4Cl (^{15}N, 99%) and or D-glucose (^{13}C, 98%). For NMR experiments the N-terminal His_6-tag was removed using TEV protease following IMAC purification. The protein was subsequently purified to homogeneity by gel filtration into reaction buffer and desalted into H_2O before being freeze-dried for storage. NMR samples were prepared at 1-2mM by reconstituting freeze-dried protein in 50mM sodium phosphate buffer pH 6.7 supplemented with 0.1mM NaN_3 and 10% D_2O.

NMR parameters. All protein NMR experiments were acquired on a Varian VNMRS 600 MHz spectrometer or a Bruker AVANCE III HD 700 MHz spectrometer equipped with cryogenically cooled detection probes. Samples were maintained at 20 °C and spectra referenced using 4,4-dimethyl-4-silapentane-1-sulfonic acid. ^1H-^{15}N HSQC experiments were recorded in 90 % H_2O/ 10% D_2O. Details of the structure determination can be found in **Supporting Information**. The chemical shifts and coordinates and have been deposited in the BMRB and PDB under accession codes 34451 and 6TG5, respectively. For titrations with MupQ, MacpD spectra were recorded at 298K and pH 8.0 whereas titrations with MupS were recorded at 288K and pH 8.0. CSPs were calculated according from the ^1H and ^{15}N chemical shift change using the minimal shift approach.[41]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Sequences of protein constructs, primers, purification data, additional functional assay data (PDF)

AUTHOR INFORMATION

Corresponding Authors

E-mail matt.crump@bristol.ac.uk ;chris.willis@bristol.ac.uk

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the BBSRC for funding (BB/R007853/1) and BBSRC/EPSRC for funding through the Bristol Centre for Synthetic Biology (BB/L01386X/1), the Bristol Chemical Synthesis Centre for Doctoral Training for P Walker (EP/L015366/1) and the doctoral training grant for A Weir. We also thank the British Commonwealth for funding for N Akter (BDCS-2017-50) and the BBSRC SWBIO DTP (M Rowe).

REFERENCES

[1] Chain, E. B., and Mellows, G. (1974) Structure of pseudomonic acid, an antibiotic from *Pseudomonas fluorescens*. *J. Chem. Soc. Chem. Commun.*, 847-848.

[2] Chain, E. B., and Mellows, G. (1977) Pseudomonic acid. 1. Structure of pseudomonic acid A - A novel antibiotic produced by *Pseudomonas fluorescens*. *J. Chem. Soc., Perkin Trans. 1*, 294-309.

[3] Chain, E. B., and Mellows, G. (1977) Pseudomonic Acid. 3. Structure of Pseudomonic Acid B. *J. Chem. Soc., Perkin Trans. 1*, 318-322.

[4] Fuller, A. T., Mellows, G., Woolford, M., Banks, G. T., Barrow, K. D., and Chain, E. B. (1971) Pseudomonic Acid - an antibiotic produced by *Pseudomonas fluorescens*. *Nature 234*, 416-417.

[5] Thomas, C. M., Hothersall, J., Willis, C. L., and Simpson, T. J. (2010) Resistance to and synthesis of the antibiotic mupirocin. *Nat. Rev. Microbiol. 8*, 281-289.

[6] Shiozawa, H., Shimada, A., and Takahashi, S. (1997) Thiomarinols D, E, F and G, new hybrid antimicrobial antibiotics produced by a marine bacterium; Isolation, structure, and antimicrobial activity. *J. Antibiot. 50*, 449-452.

[7] Shiozawa, H., Kagasaki, T., Torikata, A., Tanaka, N., Fujimoto, K., Hata, T., Furukawa, Y., and Takahashi, S. (1995) Thiomarinols B and C, new antimicrobial antibiotics produced by a marine bacterium. *J. Antibiot. 48*, 907-909.

[8] Shiozawa, H., Kagasaki, T., Kinoshita, T., Haruyama, H., Domon, H., Utsui, Y., Kodama, K., and Takahashi, S. (1993) Thiomarinol, a new hybrid antimicrobial antibiotic produced by a marine bacterium; Fermentation, isolation, structure and antimicrobial activity. *J. Antibiot. 46*, 1834-1842.

[9] Helfrich, E. J. N., and Piel, J. (2016) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep. 33*, 231-316.

[10] El-Sayed, A. K., Hothersall, J., Cooper, S. M., Stephens, E., Simpson, T. J., and Thomas, C. M. (2003) Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol. 10*, 419-430.

[11] Fukuda, D., Haines, A. S., Song, Z. S., Murphy, A. C., Hothersall, J., Stephens, E. R., Gurney, R., Cox, R. J., Crosby, J., Willis, C. L., Simpson, T. J., and Thomas, C. M. (2011) A natural plasmid uniquely encodes two biosynthetic pathways creating a potent anti-MRSA antibiotic. *PLoS One 6*, 1-9.

[12] Gao, S. S., Hothersall, J., Wu, J. E., Murphy, A. C., Song, Z. S., Stephens, E. R., Thomas, C. M., Crump, M. P., Cox, R. J., Simpson, T. J., and Willis, C. L. (2014) Biosynthesis of mupirocin by *Pseudomonas fluorescens* NCIMB 10586 involves parallel pathways. *J. Am. Chem. Soc. 136*, 5501-5507.

[13] Feline, T. C., Jones, R. B., Mellows, G., and Phillips, L. (1977) Pseudomonic acid. Part 2. Biosynthesis of pseudomonic acid A. *J. Chem. Soc., Perkin Trans. 1*, 309-318.

[14] Martin, F. M., and Simpson, T. J. (1989) Biosynthetic-studies on pseudomonic acid (mupirocin), a novel antibiotic metabolite of *Pseudomonas fluorescens*. *J. Chem. Soc., Perkin Trans. 1*, 207-209.

[15] Murphy, A. C., Gao, S. S., Han, L. C., Carobene, S., Fukuda, D., Song, Z. S., Hothersall, J., Cox, R. J., Crosby, J., Crump, M. P., Thomas, C. M., Willis, C. L., and Simpson, T. J. (2014) Biosynthesis of thiomarinol A and related metabolites of Pseudoalteromonas sp SANK 73390. *Chem. Sci. 5*, 397-402.

[16] D'Ambrosio, H. K., and Derbyshire, E. R. (2020) Investigating the role of class I adenylate-forming enzymes in natural product biosynthesis. *ACS Chem. Biol. 15*, 17-27.

[17] Schmelz, S., and Naismith, J. H. (2009) Adenylate-forming enzymes. *Curr. Opin. Struct. Biol. 19*, 666-671.

[18] Estrada, P., Manandhar, M., Dong, S. H., Deveryshetty, J., Agarwal, V., Cronan, J. E., and Nair, S. K. (2017) The pimeloyl-CoA synthetase BioW defines a new fold for adenylate-forming enzymes. *Nat. Chem. Biol. 13*, 668-674.

[19] Wang, M. L., Moynie, L., Harrison, P. J., Kelly, V., Piper, A., Naismith, J. H., and Campopiano, D. J. (2017) Using the pimeloyl-CoA synthetase adenylation fold to synthesize fatty acid thioesters. *Nat. Chem. Biol. 13*, 660-667.

[20] Chen, X. H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A. W., Gottschalk, G., Sussmuth, R. D., and Borriss, R. (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol. 188*, 4024-4036.

[21] Wu, J., Cooper, S. M., Cox, R. J., Crosby, J., Crump, M. P., Hothersall, J., Simpson, T. J., Thomas, C. M., and Willis, C. L. (2007) Mupirocin H, a novel metabolite resulting from mutation of the HMG-CoA synthase analogue, mupH in *Pseudomonas fluorescens*. *Chem. Commun. 20*, 2040-2042.

[22] Wu, J. e., Hothersall, J., Mazzetti, C., O'Connell, Y., Shields, J. A., Rahman, A. S., Cox, R. J., Crosby, J., Simpson, T. J., Thomas, C. M., and Willis, C. L. (2008) In vivo mutational analysis of the mupirocin gene cluster reveals labile points in the biosynthetic pathway: the "Leaky Hosepipe" mechanism. *ChemBioChem 9*, 1500-1508.

[23] Gulick, A. M. (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol. 4*, 811-827.

[24] Haines, A. S., Dong, X., Song, Z., Farmer, R., Williams, C., Hothersall, J., Ploskon, E., Wattana-amorn, P., Stephens, E. R., Yamada, E., Gurney, R., Takebayashi, Y., Masschelein, J., Cox, R. J., Lavigne, R., Willis, C. L., Simpson, T. J., Crosby, J., Winn, P. J., Thomas, C. M., and Crump, M. P. (2013) A conserved motif flags acyl carrier proteins for β-branching in polyketide synthesis. *Nat. Chem. Biol. 9*, 685-692.

[25] Arthur, C. J., Szafranska, A., Evans, S. E., Findlow, S. C., Burston, S. G., Philip Owen, Clark-Lewis, I., Simpson, T. J., Crosby, J., and Crump, M. P. (2005) Self-malonylation is an intrinsic property of a chemically synthesized type II polyketide synthase acyl carrier protein. *Biochemistry 44*, 15414-15421.

[26] Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T., and Dorrestein, P. C. (2006) Convergence of isoprene and polyketide biosynthetic machinery: Isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA 103*, 8977-8982.

[27] Shields, J. A., Rahman, A. S., Arthur, C. J., Crosby, J., Hothersall, J., Simpson, T. J., and Thomas, C. M. (2010) Phosphopantetheinylation and specificity of acyl carrier proteins in the mupirocin biosynthetic cluster. *ChemBioChem 11*, 248-255.

[28] Mattheus, W., Gao, L.-J., Herdewijn, P., Landuyt, B., Verhaegen, J., Masschelein, J., Volckaert, G., and Lavigne, R. (2010) Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of Its biosynthesis gene cluster. *Chem. Biol. 17*, 149-159.

[29] Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol. 6*, 493-505.

[30] Sussmuth, R. D., and Mainz, A. (2017) Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed. 56*, 3770-3821.

[31] Clark, L., Leatherby, D., Krilich, E., Ropelewski, A. J., and Perozich, J. (2018) *In silico* analysis of class I adenylate-forming enzymes reveals family and group-specific conservations. *PLoS One 13*.

[32] Hemmerling, F., Lebe, K. E., Wunderlich, J., and Hahn, F. (2018) An Unusual Fatty Acyl:Adenylate Ligase (FAAL)-Acyl Carrier Protein (ACP) Didomain in Ambruticin Biosynthesis. *ChemBioChem 19*, 1006-1011.

[33] Chang, K. H., Xiang, H., and Dunaway-Mariano, D. (1997) Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: A site-directed mutagenesis study with the Pseudomonas sp. strain CBS3 4-chlorobenzoate : coenzyme A ligase. *Biochemistry 36*, 15650-15659.

[34] Yang, J. Y., Yan, R. X., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods 12*, 7-8.

[35] Kochan, G., Pilka, E. S., von Delft, F., Oppermann, U., and Yue, W. W. (2009) Structural Snapshots for the Conformation-dependent Catalysis by Human Medium-chain Acyl-coenzyme A Synthetase ACSM2A. *J. Mol. Biol. 388*, 997-1008.

[36] Hugler, M., Menendez, C., Schagger, H., and Fuchs, G. (2002) Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation. *J. Bacteriol. 184*, 2404-+.

[37] Man, H., Kedziora, K., Kulig, J., Frank, A., Lavandera, I., Gotor-Fernandez, V., Rother, D., Hart, S., Turkenburg, J. P., and Grogan, G. (2014) Structures of Alcohol Dehydrogenases from Ralstonia and Sphingobium spp. Reveal the Molecular Basis for Their Recognition of 'Bulky-Bulky' Ketones. *Top. Catal. 57*, 356-365.

[38] Connolly, J. A., Wilson, A., Macioszek, M., Song, Z. S., Wang, L. Y., Mohammad, H. H., Yadav, M., di Martino, M., Miller, C. E., Hothersall, J., Haines, A. S., Stephens, E. R., Crump, M. P., Willis, C. L., Simpson, T. J., Winn, P. J., and Thomas, C. M. (2019) Defining the genes for the final steps in biosynthesis of the complex polyketide antibiotic mupirocin by Pseudomonas fluorescens NCIMB10586. *Sci. Rep. 9*.

[39] Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., and Owens, R. J. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications, In *Nucleic Acids Res.*

[40] Walker, P. D., Williams, C., Weir, A. N. M., Wang, L. Y., Crosby, J., Race, P. R., Simpson, T. J., Willis, C. L., and Crump, M. P. (2019) Control of beta-branching in kalimantacin biosynthesis: Application of C-13 NMR to polyketide programming. *Angew. Chem. Int. Ed. 58*, 12446-12450.

[41] Pellecchia, M., Sebbel, P., Hermanns, U., Wüthrich, K., and Glockshuber, R. (1999) Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR. *Nat. Struct. Biol. 6*, 336-339.